D23BN3 Molecular Nutrition Coursework 1:

Application of Bioinformatics to the analysis of gene expression.

**Student ID: 4248976**

### Questions (marks in brackets).

1. **Insulin and glycaemic clamps**
2. Briefly outline why somatostatin used in the clamps? (5)

Somatostatin is a hormone that inhibits the secretion of several hormones, but for this experiment, most importantly it inhibits the secretion of insulin.

1. Which of the graphs below best describes the six subjects in the low insulinemic-euglycemic (LIEu) clamp treatment group and six subjects in the low insulinemic-hyperglycaemic (LIHyp) clamp treatment group? (continued on next page) (4)

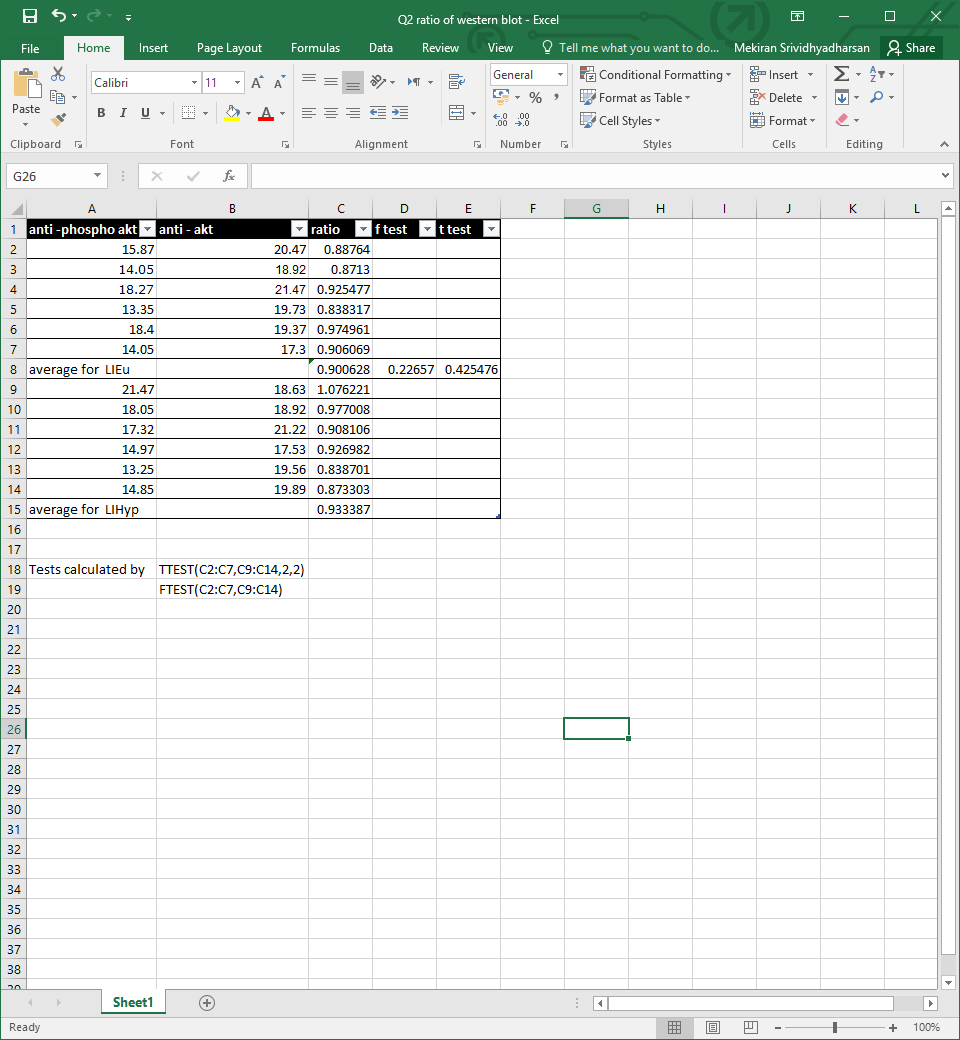
Low insulinemic-euglycemic (LIEu) clamp = B

Low insulinemic-hyperglycaemic (LIHyp) clamp = D

1. **Western blotting**
2. Using the western blot data obtained from the experiment described in this schedule is there an effect of high plasma glucose on Akt activity? (**10**)

* *To do this analysis you need to determine if the quantity of phosphorylated protein has changed relative to the non-phosphorylated protein. Use statistics to make the comparison*

The ratio was calculated by finding the average between anti-phosphorylated Akt and anti-Akt, and then dividing by the total anti-Akt. The average ratio for LIEu is 0.900628 and the average ratio for LIHyp is 0.933387. To test if there is a significant difference between the means of two groups, the t test value for this data would need to be calculated. We already know the data is two tailed as there are two samples of data, one group infused with LIEu clamp and another group infused with LIHyp clamp. To find the variance of the data, we need to calculate the f test value. The f test value is 0.22657, and since this value is more than 0.05, the variance is equal. The t test value is 0.425476 and since the value is more than 0.05, there is no statistically significant difference between the two samples. We can conclude that high plasma glucose has no effect on Akt activity, which is what we expect since insulin stimulates the phosphorylation of Akt and the clamps used both keep insulin at the same (low) level.

****

1. The objective of the experiment is to examine the effects of glucose on gene expression *in vivo*. What is your interpretation of the data obtained from the western blot? (**10**)

- *Your answer should consider why the assessment of the phosphorylation of AKT is being used a control in this experiment? What would expect to happen in this experiment? What is the hypothesis for the effects on AKT phosphorylation?*

Akt is part of the insulin signalling pathway. As a kinase it is activated by phosphorylation, so the phosphorylation of Akt is relative to the total AKT. A is a blot of Akt and B is a plot of the phosphorylated – Akt. The protein is run across the gel and the antibody will be specific to Akt and phosphorylated-Akt. A secondary advantage of this western blot is that it acts as further determination that the Akt is present since we know the size of Akt already. Stimulation of the insulin signalling pathway causes the phosphorylation of Akt, but the clamps should be keeping insulin at the same (low) level. We should expect that Akt will not be affected by changes in the glucose level. So this experiment acts as a control to see that the clamps work. From 2a, we have concluded that there is no significant difference between the Akt in both clamps, so the intensity of the bands are the same in the western blot, so glucose has no effect on Akt activity.

1. **Retrieving a sequence from the database using an accession number.**
2. Define what an accession number is? (**1**)

Accession number is a unique number assigned to a biological polymer sequence such as DNA or protein, when it is submitted to sequence databases.

1. Define what a cDNA is? (**1**)

cDNA stands for complementary DNA and it is DNA produced from a single stranded RNA template like mRNA (messenger RNA) or microRNA, that has been catalysed by the enzyme reverse transcriptase. The reaction is known as reverse transcription.

1. What is an EST, define the abbreviation and describe what they are? (**2**)

EST stands for Expression Sequence Tag and they are short pieces of sequences that are used to identify their respective expressed gene.

1. **Determining what the EST cDNA encodes for.**

a. The accession number for a sequence shown below should be within the list of sequences which have an E value = 0. What is the length of the sequence and what is its DEFINITION? (**2**)

The accession number is NM\_023085.3. The sequence has a length of 2197bp and its definition is Homo sapiens calpain 10 (CAPN10), transcript variant and mRNA.

1. What is the name of the protein that the mRNA represented by the EST encodes for and what type of enzyme is this protein likely to be? (**2**)

The proteins translated by the mRNA are calpain-10 proteins and they are calcium dependent cysteine proteases.

**5. Determine which tissues this gene product is likely to be expressed in and the number of isoforms.**

1. From your examination of the UniGene “Expression Profile”, what is the level of expression of mRNA of interest in muscle and pancreas and of all the tissues listed which has the highest expression, indicate your value as Transcripts per million (TPM). (**3**)

Muscle = 37 TPM; Pancreas = 23 TPM. Out of all the tissues, lymph has the highest expression with a value of 203 TPM.

1. From your search of the protein database (UniProt), how many isoforms of this protein are there in humans and what is the mechanism by which they are believed to be generated? How many amino acids are encoded by isoform A and isoform C and what is the predicted molecular mass of isoform D (unprocessed precursor)? (**5**)

There are 8 isoforms of the protein that are produced by alternative splicing. Isoform A is encoded by 672 amino acids and Isoform C is encoded by 517 amino acids. The predicted molecular mass of isoform D is 57816 Da.

**6. Multiple sequence alignment**

1. What is the length of AF089088 and what is its definition (DE)? (**1**)

The length of AF089088 is 2620bp and its DE is Homo sapiens calpain-like protease CAPN10a mRNA, complete cds.

1. What is the length of AF089091 and what is its definition (DE)? (**1**)

The length of AF089091 is 2001bp and its DE is Homo sapiens calpain-like protease CAPN10c mRNA, complete cds.

1. How many nucleotides would have to be deleted from the **middle** of AF089088 to give AF089091 and after what nucleotide number in AF089088 would this take place? (**2**)

1920 – 1445 = 475 nucleotides would have to be deleted

It would happened after nucleotide number 1445 in AF089088.

1. What is the reason for the absence for this sequence, identified in 6c, in AF089091? (**2**)

The absence of this sequence can be explained via alternative splicing creating alternative transcripts, so the exon in the genomic DNA that corresponds to that specific middle sequence, can be found in the CAPN10a mRNA transcript but not the CAPN10c mRNA transcript.

**7. Quantitative RT-PCR analysis of gene expression. Complete the tables below.**

Using the data in Table 2 plot a standard curve using the log cDNA weight (x axis) vs the Ct value (y axis) for each mRNA. This standard curve can be used to convert the experimental Ct values into quantities (fg) of mRNA

From the standard curves determine the following values using the equation of the line y=mx+c, where y is equivalent to Ct value. Rearrange the equation so that x (log cDNA) can be determined from Ct values.

|  |  |  |  |
| --- | --- | --- | --- |
| Gene mRNA | Gradient (m) | Crossing point (c) | Rearranged equation |
| AF089088 | -3.5351 | 29.328 | X=(Y-29.328)/-3.5351 |
| Cyclophilin | -3.3232 | 27.967 | X=(Y-27.967)/-3.3232 |
| Beta actin | -3.7606 | 28.976 | X=(Y-28.976)/-3.7606 |

First determine which reference mRNA should be used.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **beta actin** | | |
|  | subject | Ct value | log (cDNA) value | mRNA fg/ug total RNA |
| low insulinemic-euglycemic (LIEu) | 1 | 22 | 1.855023 | 71.61816 |
| 3 | 21.9 | 1.881615 | 76.14031 |
| 5 | 21.2 | 2.067755 | 116.884 |
| 7 | 21 | 2.120938 | 132.1107 |
| 9 | 21.7 | 1.934798 | 86.05927 |
| 11 | 21.7 | 1.934798 | 86.05927 |
|  |  | AVERAGE | 1.965821 | 94.81196 |
|  |  | SD | 0.105588 | 24.15855 |
|  |  |  |  |  |
| low insulinemic-hyperglycaemic (LIHyp) | 2 | 21.1 | 2.094347 | 124.2644 |
| 4 | 21 | 2.120938 | 132.1107 |
| 6 | 21.3 | 2.041164 | 109.942 |
| 8 | 21.3 | 2.041164 | 109.942 |
| 10 | 20.9 | 2.14753 | 140.4526 |
| 12 | 21.3 | 2.041164 | 109.942 |
|  |  | AVERAGE | 2.081051 | 121.1089 |
|  |  | SD | 0.046819 | 1.113831 |
|  |  |  |  |  |
|  |  | Test for variance |  | 0.213993 |
|  |  | Test for statistical significance |  | 0.041528 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **cyclophilin A** | | |
|  | subject | Ct value | log (cDNA) value | mRNA fg/ug total RNA |
| low insulinemic-euglycemic (LIEu) | 1 | 20.7 | 2.186747713 | 153.7261 |
| 3 | 20.8 | 2.156656235 | 143.4354 |
| 5 | 20.5 | 2.246930669 | 176.5756 |
| 7 | 20.6 | 2.216839191 | 164.7552 |
| 9 | 20.9 | 2.126564757 | 133.8335 |
| 11 | 21.1 | 2.066381801 | 116.515 |
|  |  | AVERAGE | 2.166686728 | 148.1401 |
|  |  | SD | 0.065005022 | 21.64627 |
|  |  |  |  |  |
| low insulinemic-hyperglycaemic (LIHyp) | 2 | 20.9 | 2.126564757 | 133.8335 |
| 4 | 20.9 | 2.126564757 | 133.8335 |
| 6 | 20.8 | 2.156656235 | 143.4354 |
| 8 | 21.2 | 2.036290323 | 108.7152 |
| 10 | 21.1 | 2.066381801 | 116.515 |
| 12 | 20.8 | 2.174121151 | 149.3211 |
|  |  | AVERAGE | 2.114429837 | 130.9423 |
|  |  | SD | 0.053018924 | 1.129845 |
|  |  |  |  |  |
|  |  | Test for variance |  | 0.48728 |
|  |  | Test for statistical significance |  | 0.145251 |

|  |  |
| --- | --- |
| What is the test for variance between the groups? | ftest |
| What is the statistical test used to compare between the two groups? | ttest |

Is there a significant difference between the two groups for each reference gene mRNA?

Cyclophilin A does not have a significant difference between the two groups, since the p value from the statistical t test is greater than 0.05, while Beta actin does have a significant difference since their p value from the statistical test is less than 0.05.

|  |  |
| --- | --- |
| Which gene should be used as the reference? | Cyclophilin A |

Since we don’t want the reference gene to be significantly different from the mRNA we want a p value above 0.05, so Cyclophilin A should be used as the reference.

Use the standard curve for AF089088 cDNA to determine the quantity of each mRNA (fg/ug total RNA) present in each low insulinemic-euglycemic (LIEu) vs low insulinemic-hyperglycaemic (LIHyp) subject.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **AF089088** | | |
|  | subject | Ct value | log (cDNA) value | mRNA fg/ug total RNA |
| low insulinemic-euglycemic (LIEu) | 1 | 28.3 | 0.290798 | 1.953431 |
| 3 | 28.9 | 0.121072 | 1.321513 |
| 5 | 29.2 | 0.036208 | 1.086947 |
| 7 | 29.4 | -0.02037 | 0.954186 |
| 9 | 29.1 | 0.064496 | 1.160102 |
| 11 | 29 | 0.092784 | 1.23818 |
|  |  |  |  |  |
| low insulinaemic-hyperglycaemic (LIHyp) | 2 | 28.7 | 0.177647 | 1.505383 |
| 4 | 27.8 | 0.432237 | 2.705433 |
| 6 | 27.5 | 0.5171 | 3.289273 |
| 8 | 28.7 | 0.177647 | 1.505383 |
| 10 | 29 | 0.092784 | 1.23818 |
| 12 | 28.5 | 0.234223 | 1.714836 |

The values in the table above need to be adjusted relative to the reference gene.

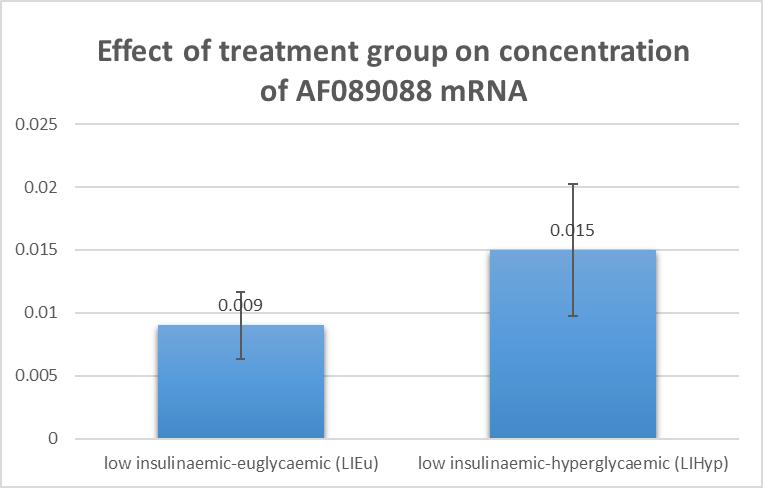
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | subject | AF089088  mRNA fg/ug total RNA | Reference mRNA fg/ug total RNA | AF089088/ reference |
| low insulinaemic-euglycaemic (LIEu) | 1 | 1.953431 | 153.7261 | 0.013 |
| 3 | 1.321513 | 143.4354 | 0.009 |
| 5 | 1.086947 | 176.5756 | 0.006 |
| 7 | 0.954186 | 164.7552 | 0.006 |
| 9 | 1.160102 | 133.8335 | 0.009 |
| 11 | 1.23818 | 116.515 | 0.013 |
|  |  |  |  |  |
| low insulinaemic-hyperglycaemic (LIHyp) | 2 | 1.505383 | 133.8335 | 0.011 |
| 4 | 2.705433 | 133.8335 | 0.02 |
| 6 | 3.289273 | 143.4354 | 0.023 |
| 8 | 1.505383 | 108.7152 | 0.014 |
| 10 | 1.23818 | 116.515 | 0.011 |
| 12 | 1.714836 | 149.3211 | 0.011 |

Using the quantitative RT-PCR data, what was the effect of the low insulinemic-hyperglycaemic clamp on the expression of AF089088 mRNA compared to the low insulinemic-euglycemic (LIEu) group?

What was the % change (increase or decrease) in the concentration of AF089088 mRNA relative to reference mRNA between the two groups?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | Test for variance  (P value) | Test for stat, signific.  (P value) | % change relative to control |
| Treatment group | mean | sd |
| low insulinemic-euglycemic (LIEu) | 0.009 | 0.002639 | 0.159001234 | 0.02694055 | **66.7%(3sf)** |
|  |  |  |
| low insulinemic-hyperglycaemic (LIHyp) | 0.015 | 0.005234 |

Show an appropriate alternative summary of the data.



Bars = mean, error bars = standard deviation

The data is significantly different since the p value is less than 0.05 (p value = 0.026941).

The percentage change relative to control is 66.7% (calculated by (0.015-0.009)/0.009 x 100).

***This section (section 7) is worth 15 marks.***

**8. Identifying a gene's position on the human genome and determination of disease states associated with the gene.**

1. What is the official symbol of the gene from which the mRNA transcripts originate (be careful the gene name is not the protein name) and what is the gene map location (position in the chromosome or cytogenetic location) of the candidate gene (**2**)

CAPN10 is the official symbol of the gene. The CAPN10 gene has been identified in a 66 kb region in chromosome band 2q37.3 within the NIDDM1 region.

1. What is the name of each gene immediately adjacent to the 5’ and 3’ ends of the gene of interest and what is their likely function? (**2**)

Adjacent to the 3’ end of CAPN10 gene is the GPR35 gene that encodes a G protein-coupled receptors, and the GPR35 is the receptor for the mucosal chemokine CXCL17. Adjacent to the 5’ end of CAPN10 gene is the RNPEPL1 gene that encodes a protein with homology to aminopeptidase B (arginyl aminopeptidase), called arginyl aminopeptidase like 1.

1. What is the disorder associated with the variable expression of this gene product? (**2**)

CAPN10 encodes a Calpain-10 protein, which is a protease that contributes to the formation of type 2 or non-insulin-dependent diabetes mellitus (NIDDM).

1. From the OMIM web pages, in what two populations was the association between the gene and disease state discovered? (**2**)

CAPN10 gene had a high-risk haplotype combination, designated 112/121, and Mexican American, Finnish and German populations were all more likely to develop type 2 diabetes due to this association. Also, CAPN10 alleles had an association with polycystic ovary disease specifically affecting Spanish and African American women.

**9. Identification of a potential promoter region with gDNA and mapping the sequence elements that potentially bind transcription factors.**

(A copy of the genomic sequence is below on which transcription factor binding elements can be marked.)

1. Why is there not a perfect alignment between the of 5’ end of the AF089088 sequence (the truncated 720bp piece used) and the genomic sequence and what is the region in the genomic DNA that has given a continuous region of alignment (i.e. no gaps)? What is the important feature that this alignment allows use to potentially define in the gene of interest *HINT: AF089088 is a cDNA sequence (a copy of mRNA) and the other is a genomic DNA sequences, so what be the reason to align them (which sequences would they have in common).* (**4**)

AF089088 is a cDNA sequence, while the other is a genomic sequence so the alignment will be between the exon in the genomic sequence and the cDNA. cDNA synthesizes from mRNA and mRNA only contains the exons from the original DNA sequence, so there are no introns. Introns in the genomic sequence is what causes the imperfect alignment.

1. What is the position of the start of transcription (initiating nucleotide) in the genomic sequence? Hint: at what nucleotide number in the gene sequence does the alignment to AF89088 cDNA start (**1**)

934

1. What basal/core transcription factor response elements in the genomic sequence indicate that this sequence may contain a promoter? This could be a combination of elements (more than one). These sequences are common to a number of genes and directly initiate “promote transcription; in this role they are position dependent. What is their position (start and end) in the genomic sequence (use the correct notation relative to initiating nucleotide) and why is this position important in terms of the sequence elements acting as a promoter? (**8**)

The TATA box is a common transcription factor that is in the promoter region of the genomic sequence. Using the candidate gene, we can figure out the position in the genomic sequence. The start is -28 and the end is -14, which matches with the typical position of a TATA box in eukaryotic genes. The TATA box is upstream of the transcription start site and this is important as during transcription a TATA binding protein binds to the TATA box region and is responsible for unwinding the double helix, allowing the gene to be transcribed.

There are two SP1 transcription factors before the transcription start site and within the promoter region, -63 to -53 and -48 to -38. This transcription factor is a GC box promoter element, and although it is in the promoter region, both are too close to the transcription start site and not in the typical position for the transcription factor to work as a promoter element.

There are three CCAAT box transcription factors, two before the transcription site at the position of -9 to -5 and -100 to -96. This is within the promoter region (~250) and the CCAAT box transcription factor is typically 60-100 bases upstream, so the sequence at -100 to -96 will work as promoter element but the sequence at -9 to -5 will not function as a promoter element as it is not the typical position.

d. What are the positions of the AP1 and CREB consensus binding sites (start and end)? What does their position relative to the transcription start site potentially indicate about their function in enabling gene transcription to take place (again use the correct notation relative to initiating nucleotide)? (**10**)

The CREB binding sites corresponds to -88 to -81, -77 to -70 and -753 to -746; the AP1 corresponds to -772 to -764, -761 to -754, and -743 to -734. The -77 to -70 and -88 to -81 CREB binding site fit in the promoter region of this genomic sequence (~250) but the CREB binding site of -753 to -746 is too far out to be a promoter. Most likely, the -753 to -746 CREB binding site and the -745 to -734 AP1 binding site are most likely enhancers or repressors, as they are located a long way from the transcription start site (~1000s bases). The AP1 sites are too far out to be promoter elements as well and AP1 transcription factors will be also most likely be enhancers or repressors to the gene of interest. An enhancers’ main function is to attract, position and modify general transcription factors and RNA polymerase II at the promoter region so that transcription begins. It may also act by modifying local chromatin structure to aid transcription. Repressors are the opposite of enhancers and they can use one or more mechanisms to repress transcription of genes.

f. The consensus sequence element 5'-YRYRGGGCACTCCYRYR-3' (where R=G or A and Y=C or T) has been identified in other genes as a hexose responsive element (HexRE), likewise 5'-CCAAAAAAGG-3' consensus sequence has been identified as the serum responsive element (SRE). Are these in this genomic sequence, if so, at what position (nucleotide numbers with correct notation) in the genomic sequence (Hint: you can use "find" in the edit drop-down menu in Word)? **(3)**

The consensus sequence element 5'-YRYRGGGCACTCCYRYR-3' is present in two sites, tacagggcactcctatg at position -810 to -793 and cacggggcactcccgtg at position -733 to -717. The consensus sequence 5'-CCAAAAAAGG-3' is present at position -788 to -779.

>candidate gene sequence 2018-19

ctgggctcctaacgaaggccctggggcccggtgggatgaaaagccctattaggtagcaaa

cagctgtcactagaggggttgaggtatggggcactcccgtgtgacgacgacggcggcagg

aggtacagggcactcctatgtgcgccaaaaaaggctgggatgtgactcacacctgactca

tgacgtcatggtgactcactcacggggcactcccgtgcccagagcccgcgtgttcaggcg

aggaaggtgcatgctgggagcggcggcgcatgctgggagctgtagtctgcgacgcaactc

ggccgaggtggctccctggtccctgaagctcccagagcccgcgtgttcaggcggtcccga

gctcccagcccagttgctgctggtggtttggcaactggctgcagagatgctgccctgaag

aattgtgctgaatgtgggtaatttagattacttgtatcttagacagtggctccctggtcg

ggaatcagattatctaacagatttctcaggttccaagccgctgaatgtggctgaatgtgg

acaaaaacctgattcctcgtaagggggcaaggcccaggattttctctatttgatgtgttc

cttccagtttttcagatggggccacgctggaagaaagaactagaagagtccagtttttcg

caccccggcccgagcctcaccggctggaggactgaacgcctgccggccctccgggtatga

ccaaaaaaggggatagccctgggacaggtgttgcccggaaggaaaaaatagggaatttcc

cgccgatcacgtgaccgggagcctatggggcactcccgtgcgacgtccactagccaatca

gtggctgacgtcatcgtgacgtcacgaagggggcgggggccacaggggcggggcccccac

gcccggtatataaagccctggttaccaatgggagactagcgggccggcgtactggcctgg

tcCAGCACCTGCGGGgccctcgggcttggagggctgggccgggcggggaacgggcggggc

gggccggaggcggcggcggctgactcgccttctctccggggctgcgaccccgaggcaacc

ggctgcagatgggagcccgcggagccgaggatgcgggcgggccggggcgcgacgccggcg

agggagctgttccgggacgccgccttccccgccgcggactcctcgctcttctgcgacttg

tctacgccgctggcccagttccgcgaggacatcacgtggaggcggccccaggtggggccg

tgtggggtgcggtgggcgccgtttctggtttctgagatctccgctcctcgcagggagcgg

ggcggggtgggcggccagggtagctccgaacgcagggtccgccgttgttctcctcagaag

tgggcgcccggccccctcttttcgtacctccttcatacccccgcccagaacgagcaggac

tcggcgctaccctaaggacgctaaactaggtcgtggcctccgcctgcgagagctccaatc

caggaggctcagafgcgctgcgagaggcgttttaacagagccccaaaaccccgccccacct

Transcription start site = +1 gactag

TATA box = -28 to -16 gtatataaagccctg

SP1 = -63 to -53 and -48 to -38 gggggcggggg

CCAAT box = -9 to -5 and -100 to – 96 ccaat

CREB = -88 to 81, -77 to -65, -753 to -746

AP1 = -772 to -764, -761 to 754, -743 to -734

Tacagggcactcctatg -810 to -793 , cacggggcactcccgtg -733 to -717

Ccaaaaaagg -788 to -779

**10. Characterizing the putative promoter of the gene using 5’ deletion analysis and site-directed mutagenesis.**

**Questions :**

1. What is the interpretation of this 5' deletion analysis of the candidate gene's putative promoter and its responsiveness to glucose? Remember to relate your interpretation to the work done on identifying the gene and the response elements within in it; the work carried out in this coursework. Try to restrict your answer to less than 500 words (**50**)

From Figure 1, the candidate sequence acts as a promoter as there is a greater expression of the reporter protein in the some of the promoter constructs created from this sequence, when compared against the negative control. There is no statistically significant difference between Prom52 and the negative control due to the same x labelling, so Prom52 is not a promoter element. The Prom113 construct has a statistically significant higher level of expression of reporter protein when compared to the negative control, based on the greater percentage activity (70% > 10%) and the difference in labelling (a instead of x), so Prom113 construct has promoter activity. Prom113 is not glucose responsive as high glucose and low glucose have the same label (a). For Prom233, there is a statistically significant higher expression of the reporter protein when compared to Prom52. More reporter protein is expressed in high glucose compared to low glucose (80% more) so Prom233 is glucose responsive. All constructs except Prom52 will have one CCAAT and two CREB transcription factors which could be responsible for the promoter activity seen in the Prom113 and Prom233.

Prom693 has a lower expression of reporter protein compared to Prom233, only being slightly higher than the negative control (d instead of x). This promoter construct has some form of repressor activity. Prom924 has the highest expression of the reporter protein in high glucose (240%) but remains the same as Prom693 for low glucose. This sequence element is most likely a glucose-responsive enhancer, based on the position and the expression of the reporter protein.

From Figure 2, the nucleotide sequence between -118 to -126 is important for the greater expression of the reporter protein, as a mutation in that sequence causes reduced promoter activity in Prom233 and Prom924 cannot work as an enhancer. From Figure 3, no constructs are significantly different to Figure 1 as the letter labelling is the same. From Figure 4 and 5, when the nucleotides position at -798 to -806 and at -837 to -845 is changed, the greater expression seen in higher glucose at Prom924 does not occur. Both nucleotide sequences are important for Prom924 to act as a glucose-responsive enhancer.

The changes caused by the mutated sequence in Figure 3 overlaps with consensus sequence element cacggggcactcccgtg at position -733 to -717. The changes caused by the mutated sequence in Figure 4 overlaps with the consensus sequence tacagggcactcctatg at position -810 to 793. Both of those are hexose responsive elements and glucose is a hexose, however since there was no reduced promoter activity in Figure 3 but there was reduced promoter activity in Figure 4, this suggests that the tacagggcactcctatg is a functional sequence that acts as an enhancer in Prom924 while cacggggcactcccgtg is not a functional sequence. The nucleotide position in figure 5 does not correspond to any transcription factors or known consensus sequence element but we can conclude that the nucleotides present at -837 to -845 are important for the enhancer to work in Prom924.

1. An antibody is available which binds to the transcription factor that is suggested to bind to HexREs (9.f) why might this be of use in this investigation? (**10**)

The binding of an antibody to the transcription factor suggested to bind in HexREs would be useful in this investigation as there is two consensus sequences in the candidate gene sequence, and Figure 3 and Figure 4 shows the effects of those sequences being changed. The antibody will be bind to the two known consensus sequences and after precipitating out the antibody bonded to the sequence, the consensus sequence can be figured out even after the mutation of the sequence in Figure 3 and Figure 4. If the antibody cannot bind the sequence is not functional as the consensus sequences has changed too much, which can act as further evidence for Figure 4. If the antibody binds to the sequence in Figure 3, this means that the mutated sequence still acts as a consensus sequence, which would mean that sequence is functional. Also, the antibody may bind to a sequence that corresponds to the activity shown in Figure 5, so somewhere in the region of -837 and -845, and after precipitating out that sequence, a new consensus sequence will be known.

**11. Abstract of findings**

## This must not be greater than **250 words, the word count MUST be specified; if word count is not included there will be a penalty of 5 marks**. (**40**)

This study focuses on the effect that glucose has on the expression of the CAPN10 gene in order to provide insight on how gene expression in skeletal muscle is affected by glucose. The effect of high glucose concentration was able to be compared by using Low insulinemic – euglycemic (LIEu) and Low insulinemic – hyperglycaemic (LIHyp) clamps and we have proven through western blotting that raising glucose concentration is independent of other factors. CAPN10 is an affected gene that encodes Calpain-10 proteins, which are calcium dependent cysteine proteases. The protein is expressed predominately in the lymph tissue and there are 8 isoforms of the protein because of alternative splicing. The protein contributes to the formation of type 2 diabetes, having a stronger effect on Mexican American, Finnish and German population. CAPN10 alleles have an association with polycystic ovary disease specifically among Spanish and African American women. There is a nucleotide sequence between -118 to -126 that is a glucose responsive transcription factor, vital for the expression of the gene. There is a functional hexose consensus sequence at -810 to -793 and a sequence between -837 to -845, which acts as glucose-responsive enhancers. The expression of CAPN10 gene is affected by glucose concentration, with a higher glucose concentration causing a greater expression of the CAPN10 gene.

**Word Count: 213**

What you did well…..

Completed all the questions – always do this as it allows you to get all the marks you can. If there is nothing submitted it can’t be marked

Very well presented work it was easy to read and see what your answers were

Some good answers showing a high level of integration of information

Good clear ability to retrieve information from the internet.

Areas to work on…..

Try to work on your data presentation skills. You will need these for presenting data in your project. Use graphs and label them appropriately. The labelling is key, especially figure legends.

It would be good if you could work on your abstract writing skills. Try to identify the most important information (data and P values) and relate this succinctly in a limited number of words. This will be needed for your dissertation but also many other instances where you need to present information.